



Boston Area Research Initiative • Northeastern & Harvard Universities
E-Mail: bari@northeastern.edu • Web: www.bostonarearesearchinitiative.net

Documentation for Cross-sectional Property Assessment Database 2024

Overview

This document describes the structure and organization of the City of Boston Assessing Department's centralized database for property-specific data for all uniquely identifiable properties in the city ($n = 182,235$) for the 2024 fiscal year. Boston's Assessing Department is responsible for determining ownership and physical characteristics for all properties in the city to ensure fair assessment of both taxable and non-taxable property in Boston. This dataset can be used to analyze valuation, structure, land use, and other details for all properties. The data are released by the Assessing Department annually as part of the City of Boston's open data initiative through *data.boston.gov*. The data are then processed by the Boston Area Research Initiative (BARI), during which additional variables are introduced to facilitate informed analysis and other aggregate measures are generated describing the properties within neighborhoods.

The main dataset (*PADCross.Record.YEAR.csv*) is the base file derived from the tax assessor's annual release through *data.boston.gov* but is curated by BARI to contain a handful of additional variables to facilitate informed analysis.

We also offer aggregated datasets that track change in assessment and use over time at the tract and block group levels for the 2010 and 2020 Census Geographies, such as the *PADCross.CT10.YEAR.csv* file. This file contains aggregate measures, also known as ecometrics, that describe neighborhoods at the 2010 census tract level. These variables are available in a spreadsheet format (*.csv*) and as mappable shapefiles (*.shp*).

Before 2016, BARI released these datasets using the name "Tax Assessor's Database" rather than "Property Assessment Database". We have changed names in order to match the names used by the City of Boston. Beginning with the 2024 release (this release), the year listed for the datasets reflects the fiscal year (in this case 2024) rather than the year when the data was collected (in this case, 2023). This change was made to reflect names by the City of Boston.



Table of Contents

1.	Summary of Record-level Property Assessment Data (PADCross.Record.YEAR.csv)	3
1.1	Description of Variables	3
1.1.1	Identifying Characteristics	3
1.1.2	Property and Building Characteristics	4
1.1.3	Geographical information	9
2.	Summary of Aggregate Measures (PADCross.CT10.YEAR.csv & corresponding shapefiles)	10
2.1	Description of Variables	10
3.	Appendix	12
3.1	Appendix A: Property Type Codes (LUC)	12
3.2	Appendix B: Codes for Land Use	13



1. Summary of Record-level Property Assessment Data (PADCross.Record.YEAR.csv)

The City of Boston's Assessing Department is responsible for determining accurate values for all properties in the city. To this end, the Department maintains property ownership and value information to ensure fair assessment of both taxable and non-taxable property in Boston. Assessing records are compiled and reviewed annually to reflect changes to properties arising from new construction, remodeling, and changes in ownership. The data contained within describes the property-specific address, ownership, type, structure, class, and valuation data. Upon annual review and re-assessment, property-specific data is updated and changes in calculated values are adjusted to reflect the most up to date status for each property. Property taxes (as indicated within this dataset as "GROSS_TAX") are also adjusted annually to reflect the annual taxation rates for residential and commercial properties.

The tax rate is the amount a taxpayer owes for each one thousand dollars of property value for that year. The tax rate for a given fiscal year appears on the third quarter tax bill, which is typically issued in late December. Commercial tax rates are calculated in the same manner.¹

This dataset is a modified version of the original Property Assessment Data provided by the City of Boston, including all variables present in the original data as well as others introduced by BARI to facilitate informed analysis. Overall, the dataset provides insight into the current physical makeup and history of properties in Boston.

1.1 Description of Variables

Property assessment variables are split into three categories: identifying characteristics, property and building characteristics, and geographical information. Identifying characteristics include variables regarding the basic identity and attributes of the address. Building characteristics include information on the physical attributes of the buildings containing the property. Geographical information provides further detail on the location of the property and the other geographies that contain it.

1.1.1 Identifying Characteristics

- *PID* is the 10-digit property identification number, unique to each property. The first two digits indicate the Ward, digits 3 through 7 are the parcel, and digits 8 through 10 are the sub-parcel.
- *CM_ID* is the 10-digit parcel number of the main condo building parcel. All condo

¹ Tax rate calculation information published by the City of Boston Assessing Department through the department's website: <http://www.cityofboston.gov/assessing/taxrates.asp>



units in each building are related to this number.

- *GIS_ID* is another 10-digit property identification number. It is the unique identifier for the land the property is in (this is slightly different from *Land_Parcel_ID*, however, as the latter combines some land parcels into one).
- *ST_NUM* is the street number of the property.
- *ST_NAME* is the street name of the property.
- *UNIT_NUM* is the specific unit number within a multi-unit building.
- *CITY* is the city where the property is located.
- *ZIPCODE* is the zip code of the property.

1.1.2 Property and Building Characteristics

- *LUC* is the land use code of the property. Properties are classified according to State Class Codes, which are three digit codes. LUC between 001 and 299 are Residential properties. LUC between 300 and 399 are Commercial properties. LUC between 400 and 465 are Industrial properties. LUC greater than 900 are Exempt properties, meaning they are fully (or partially) exempt from property taxes. See Appendix A for a full list of land use codes.
 - Note: *LUC* replaces the previous *PTYPE* column.
- *LU* denotes the type of property. Property types can be found in Appendix B.
- *OWN_OCC* is a one-character code that indicates if an owner receives a residential exemption for the property. *Y* indicates that the owner claims to live within the property (a.k.a. the property is "owner-occupied") and a *N* indicates the opposite.
- *OWNER* is the primary owner of the property, as of the date of the prior calendar year.
- *MAIL_ADDRESSEE* is the name associated with the street mailing address of the owner, if it is not the owner.
- *MAIL_ADDRESS* is the street mailing address of the owner, to which the property tax bill is mailed.
- *MAIL_CITY* is the city to which the property tax bill is mailed.
- *MAIL_STATE* is the state to which the property tax bill is mailed.
- *MAIL_ZIPCODE* is the zip code of the property where the tax bill is mailed.
- *LAND_VALUE* is the assessed value of the land.



- *BLDG_VALUE* is the total assessed value for the building on the property.
- *TOTAL_VALUE* is the total assessed value for the property. It is a summation of the assessed values of the land and building.
- *GROSS_TAX* is the amount billed to the owner as property excise tax. It is based on the total assessed value multiplied by the tax rate. Tax rates are adjusted each year for Residential and Commercial property types.
- *LAND_SF* is the total size of the property in square feet. This is also known as the lot size.
- *YR_BUILT* is the year in which the property was built. The original dataset held many properties whose year of construction was listed as zero. It was fixed by updating the *YR_BUILT* variable, which now contains a "NA" value where it previously showed a "0".
- *YR_REMODEL* is the year in which the property was last remodeled. For some properties the year of its most recent remodel was listed as zero. It was fixed by updating the *YR_REMODEL* variable, which now contains a "NA" value where it previously showed a "0".
- *GROSS_AREA* is the gross floor area for commercial properties
- *RES_FLOOR* is the number of levels in the structure that is located on the property.
- *STRUCTURE_CLASS* is the structural classification of buildings. Classes include: *A* for Steel Structure, *B* for Reinforced Concrete, *C* for Brick & Concrete, *D* for Wood/Frame, and *E* for Metal.
- *ROOF_STRUCTURE* is the roof type for structures. Types include: *F* for Flat, *G* for Gable, *H* for Hip, *L* for Gambrel, *M* for Mansard, *O* for Other, and *S* for Shed.
- *ROOF_COVER* is the exterior finish type for the roof of the properties. Types include: *A* for Asphalt Shingle, *C* for Composition, *R* for Rubber Roof, *S* for Slate, *T* for Tile, *O* for Other and *W* for Wood Shingle.
- *BED_RMS* is the total number of bedrooms in a structure.
- *FULL_BTH* is the total number of Full Bathrooms in a structure. A full bath is also known as a four-piece bath, which includes a shower, a tub, a sink, and a toilet.
- *HLF_BTH* is the number of Half Bathrooms in a structure. A half bath typically includes a sink and a toilet, also known as a powder-room.
- *BTHRM_STYLE1* is the bath style of the first bathroom in a structure. Types include: *L* for Luxury, *M* for Modern, *N* for No Remodeling, and *S* for Semi-



Modern.

- *BTHRM_STYLE2* is the bath style of the second bathroom in a structure. Types are above.
- *BTHRM_STYLE3* is the bath style of the third bathroom in a structure. Types are above.
- *KITCHEN_TYPE* is the type of kitchen in a condominium unit. Types include: F - Full Eat In, 1F - 1 Full Eat In Kitchens, 2F - 2 Full Eat In Kitchens, 3F - 3 Full Eat In Kitchens, 4F - 4 Full Eat In Kitchens, 0F - 0 Full Eat In Kitchens, N - None, O - One-Person, and P - Pullman.
- *KITCHEN_STYLE1* is the kitchen style of the first kitchen in a structure. Types include: *L* for Luxury, *M* for Modern, *N* for No Remodeling, and *S* for Semi-Modern.
- *KITCHEN_STYLE2* is the kitchen style of the second kitchen in a structure. Types are same as *KITCHEN_STYLE1*.
- *KITCHEN_STYLE3* is the kitchen style of the third kitchen in a structure. Types are same as *KITCHEN_STYLE1*.
- *HEAT_TYPE* is the type of heating in a structure. Types include: *N* for None, *S* for Space Heater, *W* for Hot Water or Steam, *E* for Electric, *P* for Heat Pump, *F* for Forced Air, and *O* for Other.
- *AC_TYPE* indicates if the structure has air conditioning. Types include: *C* for Central Air Conditioning, *D* for Ductless Air Conditioning, *Y* for Yes and *N* for None.
- *FIRE_PLACE* is the total number of fireplaces in a structure.
- *EXT_COND* is the exterior condition of a structure. Types include: *A* for Average, *E* for Excellent, *F* for Fair, *G* for Good, and *P* for Poor.
- *OVERALL_COND* is the overall condition for a structure. Types include: *VG* for Very Good, *US* for Unsound, *A* for Average, *EX* and *E* for Excellent, *F* for Fair, *G* for Good, *P* for Poor, *VP* for Very Poor, *AVG* for Default or Average.
- *INT_COND* is the interior condition of a structure. Types include: *A* for Average, *E* for Excellent, *F* for Fair, *G* for Good, and *P* for Poor.
- *PROP_VIEW* is the view for a structure. Types include: *A* for Average, *E* for Excellent, *F* for Fair, *G* for Good, and *P* for Poor, and *S* for Special.



- *NUM_BLDGS* is the number of buildings in a property.
- *BLDG_TYPE* is the building style for properties.
- *RES_UNITS* is the number of residential units in a property.
- *COM_UNITS* is the number of commercial units in a property.
- *EXT_FINISHED* is the exterior siding material for condominium buildings. Types include: A for Asbestos, B for Brick/Stone, C for Cement Board, F for Frame/Clapboard, G for Glass, K for Concrete, M for Vinyl, O for Other, P for Asphalt, S for Stucco, U for Aluminum Sliding, V for Brick/Stone veneer, W for Wood Shake, 01 for Brick, 02 for Stone, 03 for Poured Concrete, 04 for Precast Concrete, 05 for Concrete and Glass, 06 for Metal/Glass, 07 for Stone/Marble, 08 for Stucco, 09 for Wood Siding, 10 for Aluminum/Vinyl, 11 for Metal Siding, 12 for Concrete Block, 13 for Br Sill/Sash, 14 for Hollow Tile, and 15 for Corrugated Siding .
- *NUM_PARKING* is the number of parking spaces associated with a property.
- *CORNER_UNIT* indicates if a condo unit is in the corner of a building. *Y* indicates Yes and *N* indicates No.
- *LIVING_AREA* is the total living area in square feet for residential properties.
- *AV_BLDG_PER_SF* expresses the assessed value of a property's building, divided by its gross floor area in square feet.
- *AV_LAND_PER_SF* is the assessed value of a property's land, divided by the total property area in square feet.
- *SIMPLIFIED_LU* is a simplified classification of the land used code (). The Property Assessment dataset includes 17 different types of land use, including 6 different codes for residential use at varying densities (single floor houses, two-floor, etc.). Having many different classifications for similar uses can result in "artificially" high scores when used to calculate the diversity in land use distribution for a given area. *SIMPLIFIED_LU* reflects land use according to a simplified classification, with eight core uses: Residential, Commercial, Condo, Mixed Residential/Commercial, Agricultural, Industrial, Tax Exempt, and Tax Exempt by the Boston Redevelopment Authority (applies to properties that are undergoing renovation projects).
- *COOL_SCORE* is a score that ranks the relative energy efficiency of each type of cooling system, where higher scores denote more energy efficiency. This score is later used to calculate an energy efficiency index (*EE_SCORE*) for residential



buildings. Buildings with central air conditioning are assigned a score of 1, ductless air conditioning is assigned a score of 2, and buildings without air conditioning are assigned a score of 3.

- *AGE_SCORE* represents residential unit energy efficiency based on the age of a building. Older buildings have lower scores, reflecting the assumption that older buildings are less energy efficient. This score is later used to calculate an energy efficiency index (*EE_SCORE*) for residential buildings. Buildings receive a score from 0 to 4 based on when they were built or remodeled: 0 for 200 or more years ago, 1 for 150–199 years ago, 2 for 100–149 years ago, 3 for 50–99 years ago, and 4 for less than 50 years ago.
- *HEAT_SCORE* is a score that ranks the relative energy efficiency of each type of heating system, where higher scores denote more energy efficiency. This score is later used to calculate an energy efficiency index (*EE_SCORE*) for residential buildings. Buildings are given a score of 0 if they have a space heater, 1 for hot water or steam, 2 for a heat pump, 3 for forced air, and 4 for electric.
- *EE_SCORE* is an aggregate variable that combines the *HEAT_SCORE*, *COOL_SCORE* and *AGE_SCORE* in weighted sum [$EE_SCORE = AGE_SCORE + 0.75 * HEAT_SCORE + 0.75 * COOL_SCORE$]. It indicates the property specific composite energy efficiency index. This variable is only defined for properties with land usage R1, R2, or R3 due to the prevalence of missing data for denser residential properties.
 - *Note:* This variable takes into account three other variables which are otherwise not included in visualization data but necessary to calculate energy efficiency scores.
- *BLDG_AGE* is the current year minus the year in which the building was most recently remodeled or the year in which it was first built if it was never remodeled.
- *LU_DESC* is the description of the type of the property. Refer the appendix for the various Land Use Description types.
- *CD_FLOOR* is the number of commercial levels in the structure that is located on the property.
- *INT_WALL* is the condition of the wall in the interior of the structure. Types include: *N* for Normal, *S* for Substandard, *G* for Good and *E* for Elaborate
- *KITCHEN* is the number of kitchens in a structure.



- *TT_RMS* is the total number of rooms in a structure.
- *BDRM_COND* is the condition of the bedroom in a structure. Types include:
 - A* for Average, *E* for Excellent, *G* for Good, *F* for Fair, and *P* for Poor.
- *HEAT_FUEL* is the type of fuel used for heating in a structure. Types include: *Y* for Self Contained, *I* for Individual Control, *N* for None and *C* for Common.

1.1.3 Geographical information

- *X* is the longitude of the property.
 - This is derived from the City of Boston's *Parcels 2024* shapefile.
- *Y* is the latitude of the property.
 - This is derived from the City of Boston's *Parcels 2024* shapefile.
- *Land_Parcel_ID* is the unique ID of the land parcel containing the property. For more information on this ID and the geography to which it corresponds, see BARI's *Geographical Infrastructure 2024*.
- *TLID* is the identifier for the segment of road containing the property.
 - This is found by subsetting the 2023 TIGER lines street segments to only those that match the street name of the property and then finding the one that is geographically closest to the property.
- *Blk_ID_10* is the 2010 Census Block ID number.
 - This is found by spatially overlaying the longitude and latitude of the property onto the Census Blocks shapefile.
- *BG_ID_10* is the 2010 Census Group ID number.
- *CT_ID_10* is the 2010 Census Tract ID number.
- *Blk_ID_20* is the 2020 Census Block ID number.
 - This is found by spatially overlaying the longitude and latitude of the property onto the Census Blocks shapefile.
- *BG_ID_20* is the 2020 Census Group ID number.
- *CT_ID_20* is the 2020 Census Tract ID number.



2. Summary of Aggregate Measures (PADCross.CT10.YEAR.csv & corresponding shapefiles)

Neighborhood-level datasets were created that describe aggregate features of neighborhood properties. Aggregations are made at the census tract or block group level for the 2010 and 2020 Census Geographies. CT10, CT20, CBG10, or CBG20 in the filename indicates whether data is aggregated to 2010 Census tracts (CT10), 2020 Census tracts (CT20), 2010 Census block groups (CBG10), or 2020 Census block groups (CBG20). Aggregate measures are provided in both standard format (.csv) and as mappable shape files (.shp). Truncated variable names for the latter format are included in parentheses following the original variable names. Variable names for shapefiles are in parentheses.

2.1 Description of Variables

- *CT_ID_10*, *CT_ID_20*, *BG_ID_10*, or *BG_ID_20*, the ID number for a specific Census geography that the data is aggregated to (see Section 1.1.3 for details).
- *EE_SCORE.res* (*EESR*) is the average energy efficiency index (*EE_SCORE* above) for residential properties in the area. For more information on how the energy efficiency index is calculated, see the documentation above.
- *AV_LAND_PER_SF.res* (*ALPSFR*) is the average assessed value of a property's land, divided by the total property area in square feet (*TOTAL_PER_SF* above) for all residential properties in the area.
- *AV_LAND_PER_SF.nonres* (*ALPSFN*) is the average assessed value of a property's land, divided by the total property area in square feet (*TOTAL_PER_SF* above) for all non-residential properties in the area.
- *AV_BLDG_PER_SF.res* (*ABPSFR*) is the mean assessed value of a property's building, divided by its gross floor area in square feet (*AV_BLDG_PER_SF* above) for all residential properties in the area.
- *AV_BLDG_PER_SF.nonres* (*ABPSFN*) is the mean assessed value of a property's building, divided by its gross floor area in square feet (*AV_BLDG_PER_SF* above) for all non-residential properties in the area.
- *YR_BUILT_REMOD.res* (*YBRR*) is the mean value of the latest year remodeled or the year built for all residential properties in the area.
- *YR_BUILT_REMOD.nonres* (*YBRN*) is the mean value of the latest year remodeled or the year built for all non-residential properties in the area.
- *DEC_BUILT_REMOD.res* (*DBRR*) is the modal value of the latest decade it was remodeled in or the decade it was built in for all residential properties in the area.



- *DEC_BUILT_REMOD.nonres (DBRN)* is the modal value of the latest decade it was remodeled in or the decade it was built in for all non-residential properties in the area.
- *nbhdval (random)* represents the residuals extracted from a multilevel linear regression model that calculated the unique effect each census tract had on property values when controlling for the lot size, gross floor area, total living area, number of floors, and age of residential buildings located within it. The top 4% of property values (*TOTAL_VALUE*) were dropped and the outcome variable was log- transformed to prevent any impacts on outliers on model results (final values were un-transformed for interpretability). Census tracts that have fewer than 7 properties were dropped from the model. Values can be interpreted as the estimated value of a single-family residential building with average size and age in a census tract, though the interpretation of the positive (negative) effect on values is extensible to all land uses. Higher values are indicative of a more positive effect of the census tract and lower values are indicative of a negative effect. This measure is not included for census block groups owing to small within-geography sample sizes.



3.2 Appendix B: Codes for Land Use

USE CODE	DESCRIPTION
A	Residential 7 or more units
AH	Agricultural/Horticultural
C	Commercial
CC	Commercial condominium
CD	Residential condominium unit
CL	Commercial land
CM	Condominium main (physical structure housing all related condo units with no assessed value)
CP	Condo parking
E	Tax-exempt
EA	Tax-exempt (121A)
I	Industrial
R1	Residential 1-family
R2	Residential 2-family
R3	Residential 3-family
R4	Residential 4 or more family <input type="checkbox"/>
RC	Mixed use (res. and comm.)
RL	Residential land